>G481 (gf=13) (nid=2398526) (acc=Y13723) 239. 2398526|emb|Y13723|ATHAPIA Arabidopsis thaliana tr Hap1a transcription factor
GAGCGTTTCGTAGAAAAATTCGATTTCTCTAAAGCCCTAAAACTAAAAGACTATCCCCA
ATTCCAAGTTCTAGGGTTTCCATCTTCCCCAATCTAGTATAAATGGCGGATACGCCTTCG
AGCCTAGCTGGAGATGGCGGAGAAAGCGGCGGTTCCGTTAGGGAGCAGGATCGATACCTT
CCTATAGCTAATATCAGCAGGATCATGAAGAAAGCGTTGCCTCCTAATGGTAAGATTGGA
AAAGATGCTAAGGATACAGTTCAGGAATGCGTCTCTGAGTTCATCAGCTTCATCACTAGC
GAGTCCAGTGATAAGTGTCAAAAAGAGAAAAGGAAAACTGTGAATGGTGATGATTTGTTG
TGGGCAATGGCAACATTAGGATTTGAGGATTACCTGGAACCTCTAAAGATATACCTAGCC
AGGTACAGGGAGTTGGAGGGTGATAATAAGGGATCAGGAAAGAGTGGAGATGGATCAAAT
AGAGATGCTGGTGGCGGTGTTTCTGGTGAAGAAATGCCGAGCTGGTAAAAGAAGTTGCAA
GTAGTGATTAAGAACAATCGCCAAATGATCAAGGGAAATTAGACATCAGTGAGTTGTTTA
TAGTTGAGCTGATCGACAACTATTTCGGGTTTACTCTCAATTTCGGTTATGTTAGTTTGA
ACGTTTGGTTTATTGTTTCCGGTTTAGTTGGTTGTATTTAAAGATTTCTCTGTTAGATGT
TGAGAACACTTGAATGAAGGAAAAATTTGTCCACATCCTGTTGTTATTTTCGATTCACTT
TCGGAATTTCATAGCTAATTTATTCTCATTTAATACCAAATCCTTAAATTAA
>P1_G481 fr=+1 (gf=13) (nid=2398526) (acc=Y13723) 2398526 gi|2398526|emb|Y13723|ATHAPIA Arabidopsis thaliana mRNA for Hap1a transcription fa
ERFVEKFDFSKALKLKPLSPIPSSRVSIFPNLV*MADTPSSPAGDGGESGGSVREQDRYL
PIANISRINKKALPPNGKIGKDAKDTVQECVSEFISFITSEASDKCQKEKRKTVNGDDLL
WAMATLGFEDYLEPLKIYLARYRELEGDNKGSGKSGDGSNRDAGGGVSGEEMPSW*KKLQ
VVIKNNRQMIKGN*KSVSCL*LS*STTISGLLSISVMLV*TFGLLPPV*LVVFKDFSVRC
*EHLNEGKICPHPVVIFDSLSEFHS*FILI*YSILKL
>M1_G481 fr=-1 (gf=13) (nid=2398526) (acc=Y13723) 2398526 gi|2398526|emb|Y13723|ATHAPIA Arabidopsis thaliana mRNA for Hap1a transcription fact
LI*GFGIF*E*IEYEIPKVNRK*QQDVDKFFLHSSVLNI*QRNL*IQPTKPETINQTFKL
T*PKLPVNPK*LSISSTINNSLISHFP*SFGDGS*SLLATSFTSSAFLHQKHAHQHLYLI
HLHCFLIPYYHPPTPCTSLGISLEVPSNPQILHLFLPTTNHHSQPSFSLFDTYIWPE**
*S**TJPKIPELYP*HLFQSYH*EATLGS*SC*Y*L*EGIDPAP*KNKKFLRHLQLGSKA
YPPFILDWGRWKP*NLELGIVVLVLGL*RNKIFLPNA
>P2_G481 fr=+2 (gf=13) (nid=2398526) (acc=Y13723) 2398526 gi|2398526|emb|Y13723|ATHAPIA Arabidopsis thaliana mRNA for Hap1a transcription fact
SVS*EKSISLKP*N*HDYPGFQVLGPPSSPI*YKWRIRLPAGLEMAEKAAVPL/NRIDTF
[illegible]
[illegible]
**LFTIAK*SKEIPDQ*VVYS*ADKLLFRVYSSFKLS*PKRLVYCFKFSWLYLKISLLAV
DHT*MKEKFVHILLLPSIHFKNFIANLFSPNTKSLN*
>M2_G481 fr=-2 (gf=13) (nid=2398526) (acc=Y13723) 2398526 gi|2398526|emb|Y13723|ATHAPIA Arabidopsis thaliana mRNA for Hap1a transcription fact
*FKDLVLNENKLAMEFRK*IFNNNEMWTNFSFIQVFSTSNKEIFKYAQLGKKQ*TKKSN*
INKN*E*TFNSCRSAGL*TTH*SLISLDHIAIVISHIYLASLLPARHFPTRNTAT[illegible]*G
[illegible]*SLIITLQLPVPR*VYL*RFQVILKS*[illegible]ITDSPPFLFLTLIPHLASD
FADELRDAFLNCILSIFSNLTIRRQRPLHDPADISYRKVSILGPNGTAAFSAISSWARPR
IRMLY*IGEDGNPRTWNAG*SF*F*GFKEIEFFYETL
>P3_G481 fr=+3 (gf=13) (nid=2398526) (acc=Y13723) 2398526 gi|2398526|emb|Y13723|ATHAPIA Arabidopsis thaliana mRNA for Hap1a transcription fact
AFRKSIPFL*SPKTKPTIPNSKF*GFHLPGSSINKSYAFSPSWKWRRKRRFR**AGSIPS
YSPYGGSHEESVAS*W*DWKRC*GYSSGMRL*VHQLHH*RGQ**VSKPEKENCEW**FVV
GNENIKI*GLPGTSKDIPSEVGGVSG***GIRKEWRWIK*RCWWRCPW*RNAELVKEVAS
SD*EGSPNDQGKLEISELFIVELIQNYFGPTLNFGYVSLNVWFIVSGLVGCI*KPLC*ML
FGLE*KKNLSTSCCYFRPTFGIS*LIYSHLIPNP*I
>M3_G481 fr=-3 (gf=13) (nid=2398526) (acc=Y13723) 2398526 gi|2398526|emb|Y13723|ATHAPIA Arabidopsis thaliana mRNA for Hap1a transcription fact
NLRWY*MRIN*L*NSESESKITTKGGQIFPSFKCSQHLTEKSLNTTN*TGNNKPNVGTN
ITEIESKPEIVVDQLNYKQLTDL*FPLIIWRLFLITTGNFFYQLGISSPETPPPASLFDP
SPLFPDPLLSPSNSLYLARYIFRGSR*SSNPNVAIAHNKSSPPTVFLFSF*HLSLASLVM
KLMNSETHS*TVSLASFPILPLQGNAFPMILLILAIGRYRSCSLTEPPLSPPSPAGLEGV
SAIYTRLGKMETLELGIGDSRFSFRALEKSNFSTKR
>G482 (gf=13) (nid=2398528) (acc=Y13724) 2398528 gi|2398528|emb|Y13724|ATHAP3B Arabidopsis thaliana mRNA for Hap3b transcription factor
CCGCGACAGGGATTCCGGTGGAGGGCAAAACGGGAACAACCAGAACGGACAGTCCTCCTT
GTCTCCAAGAGAGCAAGACAGGTTCTTGCCGATCGCTAACGTCAGCCCGATCATGAAGAA
GGCTTTGCCCGCCAACGCCAAGATCTCTAAAGATGCCAAAGAGACGATGCAGGAGTGTGT
CTCCCAGTTCATCAGCTTCGTCACCGGAGAAGCATCTGATAAGTGTCAGAAGGAGAAGAG
GAAGACGATCAACGGAGACGATTTGCTCTGGGCTATGACTACTCTAGGTTTTGAGGATTA
TGTTGAGCCATTGAAAGTTTACTTGCAGAGGTTTAGGGAGATCGAAGGGGAGAGGACTGG
AGTAGGGAGGCCACAGACTGGTGGTGAGGTCGGAGAGCATCAGAGAGATGCTGTCGGAGA
TGGCGGTGGGTTCTACGGTGGTGGTGGTGGGATGCAGTATCACCAACATCATCAGTTTCT
TCAGCAGCAGAACCATATGTATGGAGCCACAGGTGGCGGTAGCGACAGTGGAGGTGGAGC
TGCCTCCGGTAGGACAAGGACTTAACAAAGATTGGTGAAGTGGATCTCTCTCTGTATATA
GATACATAAATACATGTATACACATGCCTATTTTTACGACCCATATAAGGTATCTATCAT
GTGATAGAACGAACATTGGTGTTGGTGATGTAAAATCAGATGTGCATTAAGGGTTTAGAT
TTTGAGGCTGTGTAAAGAAGATCAAGTGTGCTTGTTGAACAATAGGATTCACTAACGAAT
CTGCTTCATTGAATCTTGTATGTAACTAAAGCCATTGTATTGAATGCAAATGTTTTCATT
TGGGATGCTTTAAAAAAAAAAAAAAAAAAAAAA
>P1_G482 fr=+1 (gf=13) (nid=2398528) (acc=Y13724) 2398528 gi|2398528|emb|Y13724|ATHAP3B Arabidopsis thaliana mRNA for Hap3b transcription factor
PRQGFRWKAKREQPERTVLLVSKRARQVLADR*KCPDHSEGLAKQRQDL*RCQRDDAGVC
LRVHQLAHRRSI**VSEGEEDDQRHRFALGYDYERF*GLC*AIESLLAEV*GDRRGEDW
TKEATDWR*GKKASERCCRRWRWVLRWWWDAVSPTSSVSSPAEPYVWSHRWR*RQWRWS
CLR*DKDLTKIGEVDLSLY!DT*IHVYTCLFLRPI*GIYHVIERTLVLVM*NQMCIKGLD
FEAV*KRSSVLVEQ*DSLTNLLH*ILYVTKAIVLNANVFIWDALKKKKKK
>M1_G482 fr=-1 (gf=13) (nid=2398528) (acc=Y13724) 2398528 gi|2398528|emb|Y13724|ATHAP3B Arabidopsis thaliana mRNA for Hap3b transcription factor
FFFFFFFLKHPK*KHLHSIQWL*LHTRFNEADSLVNPIVQQAHLIFFTQPQNLNP*CTSD
FTSPTPMFVLSHDRYLIHVVKIGMCIHVFMYLYTERDPLHQSLLSPCPTGGSSTSTVATA
TCGSIEKVLLVKKLMDLVILHPTTTTVEPTAISDSISLMLSDLTTSLWPP*SSPLPYSLP
KPLQVNFQWLNIILXT*SSHSPEQIVSVDRLPLLLTLIRCFSGDEADELGDTLLHRLFG
IFRDLGVGGQGLLHDPADVSDRQEPVLLSWRQGGLSVLVVPVLPSTGIPVA
>P2_G482 fr=+2 (gf=13) (nid=2398528) (acc=Y13724) 2398528 gi|2398528|emb|Y13724|ATHAP3B Arabidopsis thaliana mRNA for Hap3b transcription factor
RDRSGGGQNGNNQNGQSSLSPREQDRFLPIANVSRIMKKALPANAKISKDAKETMQECV
SEFISFVTGEASDKCQKEKRKTINGDDLLWAMTTLGFEDYVEPLKVYLQRFREIEGERTG
LGRPQTGGEVGEHQRDAVGDGGGFYGGGGGMQYHQHHQFLHQQNHMYGATGGGSDSGGGA
ASGRTRT*QKLVKWISLCI*IHKYMYTHAYFYDPYKVSIM**NEHWCW*CKISALRV*I
LRLCKEDQVCLLNNRIH*KICFIESCM*LKPLY*MQMFSFGML*KKKKKK
>M2_G482 fr=-2 (gf=13) (nid=2398528) (acc=Y13724) 2398528 gi|2398528|emb|Y13724|ATHAP3B Arabidopsis thaliana mRNA for Hap3b transcription factor
FFFFFFF*SIPNENICIQYNGFSYIQDSHKQIR**ILLFNKHT*SSLHSLKI*CLNAHLI
LHHQHGGSFYHHIDTLYGS*K*ACVYMYLCIYIQREIHPTNLS*VLVLPEAAPPPLSLPP
PVAPYIHFCW*RN**CW*YCIPPPPP*NPPPSPTASL*GSPTSPPVGGLPSPVLSPSISL
NLCH*TENSST*SSKPRVVIAQSKSSPLIVPLFSF*HLSDASPVTKLMNSETHSCIVSLA
KLSILAIASKAPFMIRLTLAIGKNLRSSLGDESSCPFWLFPFSPPPESLSR
>P3_G482 fr=+3 (gf=13) (nid=2398528) (acc=Y13724) 2398528 gi|2398528|emb|Y13724|ATHAP3B Arabidopsis thaliana mRNA for Hap3b transcription factor
ATGIPVEGKTGTTRTDSPPCLQESKTGSCRSLTSASS*RRPCPPTPRSLEMPKRKCRSVS
PSSSASSPEKHLISVRRRRGRRSTETICSGL*LL*VLKIMLSH*KFTCRGLGRSKGRGLD
*GGRRLVVRSESIREMLSEMAVGSTVVVVGGSITNIISFFTSRTICMEPQVAVATVEVEL
PPVGQGLNKDW*SDSLSVYRYINTCIHMPIFTTHIAYLSCDRTNIGVGDVKSDVH*GFRF
*GCVKKIKCAC*TIGFTNESASLNLVCN*SHCIECKCFHLGCFKKKKKK
>M3_G482 fr=-3 (gf=13) (nid=2398528) (acc=Y13724) 2398528 gi|2398528|emb|Y13724|ATHAP3B Arabidopsis thaliana mRNA for Hap3b transcription factor
FFFFFFFFKASQMKTFAFNTNALVTYKIQ*SKFVSENYCSTSTLDLLYTASKSKPLMHI*F
YITNTNVRSIT**IPYMGRKNKHVYTCIYVSIYKERSTSPIFVKSLSYRKQLHLHCKYRH
LWLHTYGSAGEETDDVGDTASHHHHRRTHHHLRQHLSDALRPHHQSVASLVQSSPLRSP*
TSASKLSMAQHNPQNLE*S*PRANRLR*SSSSSPSDTYDMLLR*RS**TRRHTPASSLWH
L*PSWKWRAKPSS*SG*R*RSARTCLALLETRRTVRSGGSAFALHRNFCR

Table 1

Original Residue	Conservative Substitutions
Ala	ser
Arg	lys
Asn	gln; his
Asp	glu
Cys	ser
Gln	asn
Glu	asp
Gly	pro
His	asn; gln
Ile	leu, val
Leu	ile; val
Lys	arg; gln; glu
Met	leu; ile
Phe	met; leu; tyr
Ser	thr
Thr	ser
Trp	tyr
Tyr	trp; phe
Val	ile; leu

An alternative indication is to show whethr two nucleic acid molecules are closely related is that the two molecules hybridize to each other under stringent conditions. Stringent conditions are sequence dependent and are different under different environmental parameters. Generally, stringent conditions are selected to be about 5°C to 20°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Conditions for nucleic acid hybridization and calculation of stringencies can be found in Sambrook et al., Molecular Cloning. A Laboratory Manual, Ed. 2, Cold Spring Harbor Laboratory Press, New York (1989)) and Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York (1993). Nucleic acid molecules that hybridize under stringent conditions will typically hybridize to a probe based on either the entire cDNA

or selected portions of the cDNA under wash conditions of 0.2x SSC, 0.1% SDS at 65°C.

For conventional hybridization the hybridization probe is conjugated with a detectable label such as a radioactive label, and the probe is preferably of at least 20 nucleotides in length. As is well known in the art, increasing the length of hybridization probes tends to give enhanced specificity. The labeled probe derived from the *Arabidopsis* DNA binding domain cDNA sequence may be hybridized to a plant cDNA or genomic library and the hybridization signal detected using means known in the art. The hybridizing colony or plaque (depending on the type of library used) is then purified and the cloned sequence contained in that colony or plaque isolated and characterized.

The degeneracy of the genetic code further widens the scope of the present invention as it enables major variations in the nucleotide sequence of a DNA molecule while maintaining the amino acid sequence of the encoded protein. For example, assume that a given amino acid residue a transcription factor protein shown is alanine, and the codon encoding this residue in the cDNA is GCA. Because of the degeneracy of the genetic code, three other nucleotide codon triplets: GCT, GCC and GCG, also code for alanine. Thus, the nucleotide sequence of the transcription factor ORF could be changed at this position to any of these three codons without affecting the amino acid composition of the encoded protein or the characteristics of the protein. Based upon the degeneracy of the genetic code, variant DNA molecules may be derived from the cDNA and gene sequences disclosed herein using standard DNA mutagenesis techniques as described above, or by synthesis of DNA sequences. Thus, this invention also encompasses nucleic acid sequences which encode a transcription factor protein but which vary from the disclosed nucleic acid sequences by virtue of the degeneracy of the genetic code.

Overall, DNA binding domains that are homologs of the disclosed DNA binding domains will typically share at least 40% nucleotide sequence identity with a DNA binding domain sequence. More closely related DNA

binding domains may share at least 50%, 60%, 65%, 70%, 75% or 80% sequence identity with the disclosed nucleotide sequences. Transcription factors that are most closely related to the disclosed nucleotide sequences share at least 85%, 90% or 95% sequence identity with one or more of the disclosed *Arabidopsis* transcription factor proteins.

Homologs of the *Arabidopsis* transcription factors may alternatively be obtained by immunoscreening of an expression library. With the provision herein of the disclosed transcription factor nucleic acid sequences, the polypeptide may be expressed and purified in a heterologous expression system (e.g., *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for the transcription factor, and preferably for the DNA binding domain. Antibodies may also be raised against synthetic peptides derived from transcription factor amino acid sequence presented herein. Methods of raising antibodies are well known in the art and are described in Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York. Such antibodies can then be used to screen an expression cDNA library produced from the plant from which it is desired to clone the transcription factor cDNA homolog, using the methods described above. The selected cDNAs can be confirmed by sequencing and DNA binding activity.

The DNA binding specificity of these polypeptides may be identified and then based on knowledge of the binding specificity of the disclosed binding domains additional DNA binding domains closely related in function may be identified. The DNA binding specificities of the recombinant polypeptides may be analyzed by gel retardation experiments. Another test to identify DNA binding specificity involves expressing the polypeptides in yeast strains which contain the target DNA regulatory sequence fused to the *lacZ* reporter gene. If the protein is a transcriptional activator, it should activate expression of the reporter gene and result in blue colonies. Yet another test involves a plant transient assay. In this case, a reporter gene, such as GUS, carrying the target DNA regulatory sequence as an upstream activator is introduced into plant cells (e.g. by particle